eVLBI and the IETF

Marshall Eubanks

TSV / IETF 67

November, 2006

What is VLBI ?

- Very Long Baseline Interferometry
- How do you make a radio telescope 10,000 km across ?
  - You connect smaller ones…
  - Historically, close telescopes are used in arrays, but data from VLBI telescopes are recorded and compared later
    - We will fix that.
- Believe it or not, there are time sensitive applications of this
  - GPS, Spacecraft Navigation, Transient Phenomenom

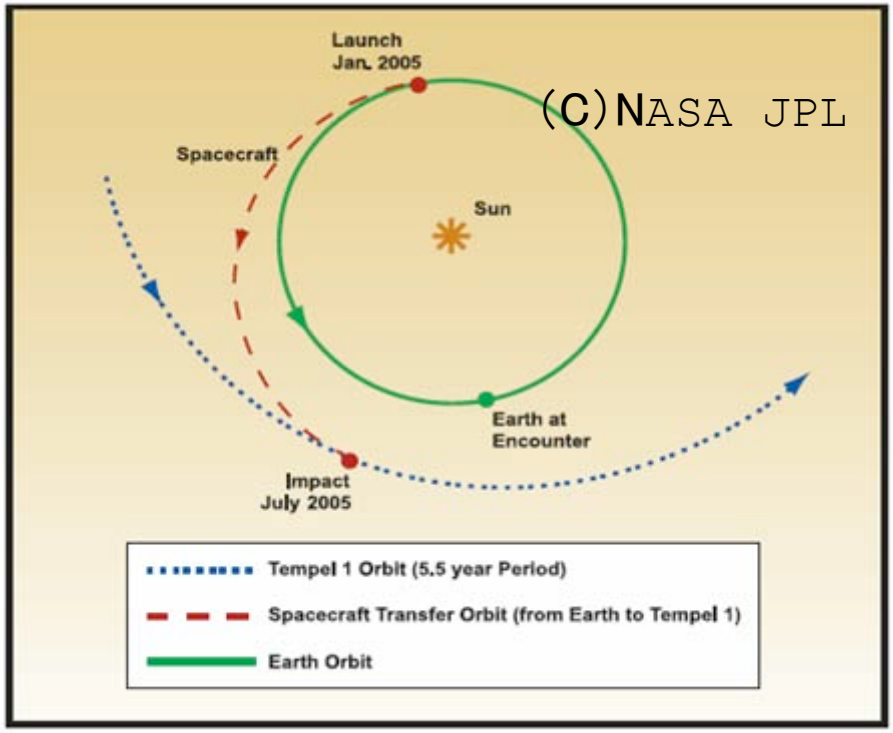
VLBI to eVLBI

- eVLBI : VLBI with electronic data transmission
- Characteristics :
  - High data rates (512 Mbps to 1 Gbps now, plans extend to 100 Gbps)
    - Can be real time, or quasi real time (transmit while the telescopes are moving) or to a buffer
  - Loss tolerant (up to ~ 1 % packet loss may be OK)
  - Each sample is typically 1 or 2 bits (so one packet contains thousands of samples)
  - Typically Many to One (Telescopes to Correlator)
- The desire is to use as much of the existing IETF infrastructure as possible
- There was a workshop on this at MIT/Haystack Observatory in late September
  - http://www.haystack.edu/geo/vlbi_td/abstract.html

Contribution to Deep Impact mission

4 July 2005

http://deepimpact.jpl.nasa.gov/



- UT1 value provided by INT session
- eVLBI observations for IVS-INT2
- Data transfer for short time
- Contribution to the success of the mission

Traveling at a relative velocity of 10 km/s and from about 864,000 km (536,865 miles) away, the impactor must strike the 6 km (3.7 mile) diameter comet.

Telescope sites participating

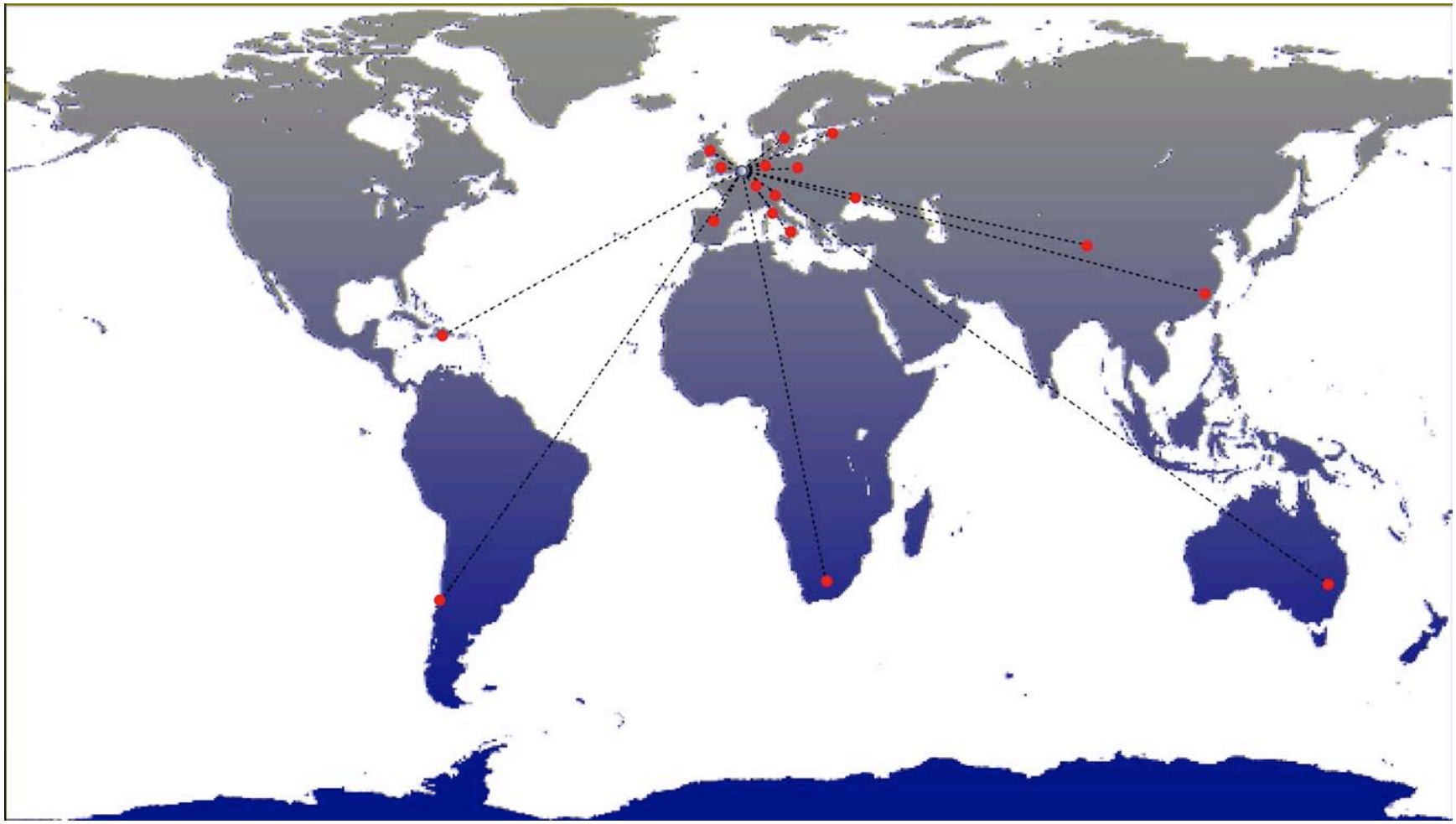
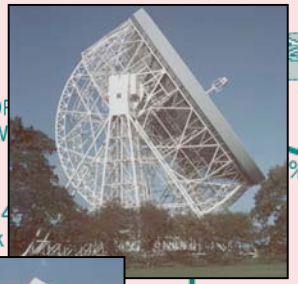
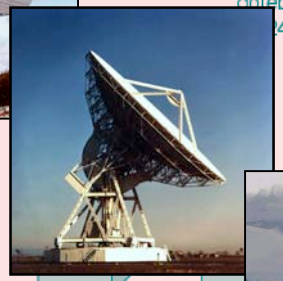
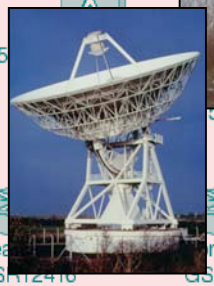


Image courtesy of Dr. Francisco Colomer,
<http://www.oan.es/expres/status.htm>

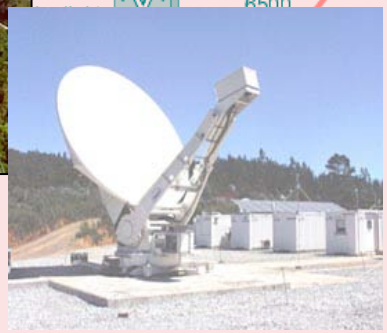
Routes across GEANT
used by eVLBI MkVs







RTT & % load

% load is approx daily high value

Cisco 3508
Cisco 6500
- Cisco 15216 EDF
- Cisco 15252 DW
- Juniper M160
Juniper T640
X - Cisco GSR 124
08 - Extreme Black

3812 - 3C
AT-9816

stockholm2
0ms ?%
0ms 20%
se-kth
T640

stockholm1
1241X
5ms 10%
jonkoping1
1241X
0.5ms 10%
boras2
1241X
2ms 10%
goteborg2
X410
0.3ms 40%
chalmers2
?
0ms ?%
oso-gw
?
?
1ms ?%
ONSALA, SE
Mk5

man'-bar
GSR12016
0.5ms 30%
Cisco1545
0.2ms 40%
ManCompCen
6500
Transmode
5400
0.5ms

36ms 5%

1 Gbps

uk1 - M160
0.2ms 30%
geant-gw3
GSR12404
0ms
5%
nd-scr3
GSR12416
7.5ms
5%
nl1 - M160
0.4ms 5%
CR1
12416
0.2ms 5%
AR5
12416
7ms
5%
cz1 - M160
20ms 5%
pl1 - M160
4m
AT-9816
TORUN, PL
160

2ms
15216

155 Mbps

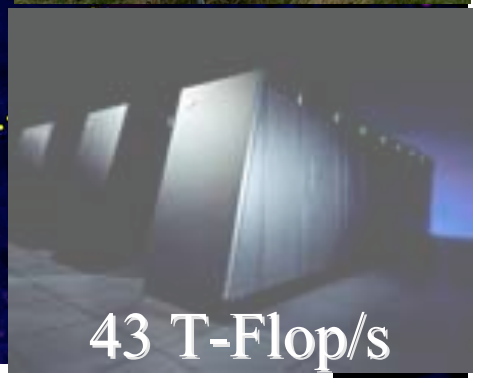
Gbps

2.5 Gbps

Bologna
M320

LOFAR

New generation radio telescope



IBM BlueGene/L processor

eVLBI data transport

- Complaints about the difficulty in realizing multi Gbps data transport.
  - "Why do **we** need to be networking experts?"

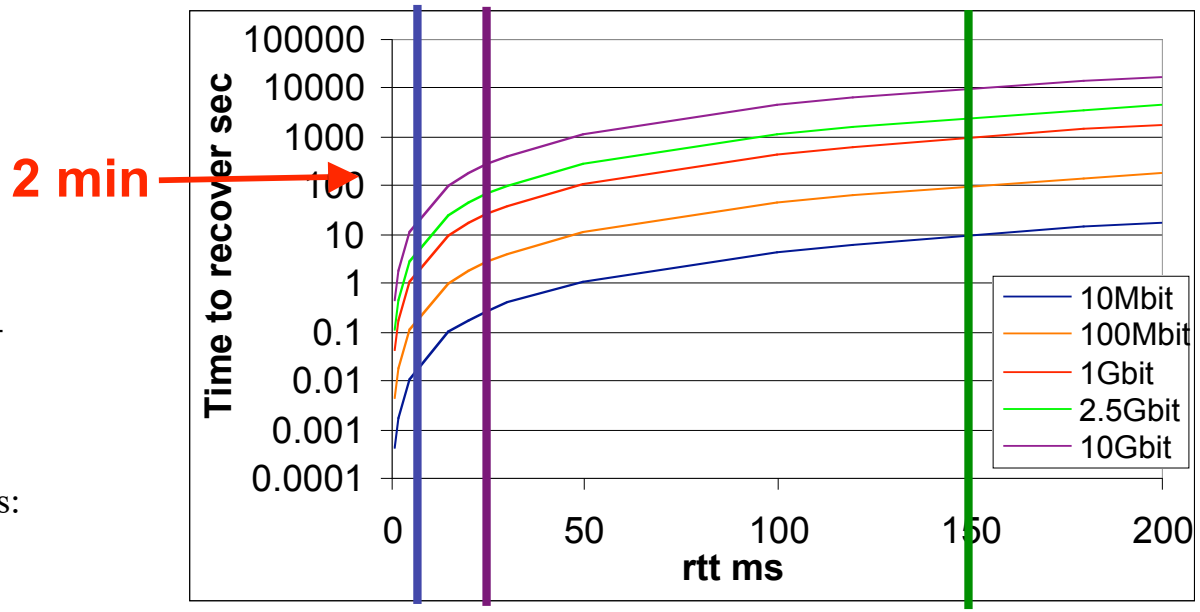
    Charles Yun, Program Manager EXPReS Project JIVE
- Responses
  - Tuning of TCP
  - New TCP Stacks
  - Use of UDP
    - Less than best effort transport
- Move to Lightpaths and GMPLS

TCP (Reno) – Details of problem #1

- Time for TCP to recover its throughput from 1 lost 1500 byte packet given by:

$$\rho = \frac{C * RTT^2}{2 * MSS}$$

- for rtt of ~200 ms @ 1 Gbit/s:



UK 6 ms Europe 25 ms USA 150 ms

1.6 s 26 s 28min

Throughput	Window	Loss recovery time	Supporting loss rate
10Mbps	170pkts	17s	5.4×10^{-5}
100Mbps	1700pkts	2mins 50s	5.4×10^{-7}
1Gbps	17000pkts	28mins	5.4×10^{-9}
10Gbps	170000pkts	4hrs 43mins	5.4×10^{-11}

(Richard Hughes-Jones The University of Manchester)

Investigation of new TCP Stacks

- **The AIMD Algorithm – Standard TCP (Reno)**
  - **For each ack in a RTT without loss:**

  cwnd -> cwnd + a / cwnd - Additive Increase, a=1
  - **For each window experiencing loss:**

  cwnd -> cwnd – b (cwnd) - Multiplicative Decrease, b= ½
- **High Speed TCP**

a and b vary depending on current cwnd using a table
  - **a increases more rapidly with larger cwnd – returns to the 'optimal' cwnd size sooner for the network path**
  - **b decreases less aggressively and, as a consequence, so does the cwnd. The effect is that there is not such a decrease in throughput.**
- **Scalable TCP**

a and b are fixed adjustments for the increase and decrease of cwnd
  - **a = 1/100 – the increase is greater than TCP Reno**
  - **b = 1/8 – the decrease on loss is less than TCP Reno**
  - **Scalable over any link speed.**
- **Fast TCP**

Uses round trip time as well as packet loss to indicate congestion with rapid convergence to fair equilibrium for throughput.
- **HSTCP-LP, H-TCP, BiC-TCP**

Enter Connection Oriented Services

- Connection Oriented Services are point to point paths set up across a network that have dedicated network resources associated with the path.
  - Example: Phone line- 64Kbps dedicated capacity between the calling party and the called party.
  - Example: OC192 SONET circuit- 9.4 Gbps between Onsala and Haystack (looking ahead☺)
- In general, connection oriented services provide a means for the user to specify service requirements for a flow, and allows the network to allocate sufficient resources to this flow apriori (before the initiating the flow) and then to release those resources when the user no longer requires them.
  - This process is called provisioning, and includes path selection and establishment at each network element along the path.
  - Often this is manual process, sometimes semi-automated,
  - Emerging experimental networks (such as DRAGON, and similar projects in Japan and Europe) are developing the tools and technologies for fully automated circuit establishement.

Courtesy Jerry Sobieski - jerrys@maxgigapop.net

“Light Paths”: Terminology for the New Millenium

- The emergence of very high capacity and low cost optical wavelength based telecommunications technologies made the prospect of dedicated and [almost] free capacity an attractive and seemingly achievable networking nirvana…
- Alas, waves are not free, or cheap…
- But they are less expensive than traditional carrier services,
- And they provide enormous capacity (10 Gbps is the norm today)
- So the concept of a wave, or “light path”, for every project that needed high capacity or predictable and repeatable performance began to take shape…
  - And it is now used to describe the new models for circuits and connection oriented services being explored in current optical networks

- A “light path” is a new term that refreshes the ideas for connection oriented services – Light paths complement IP services, and are generally integrated with IP networks, and yet promote the proposition that dedicated, predictable, and repeatable network services are required even today with such high performance networks.

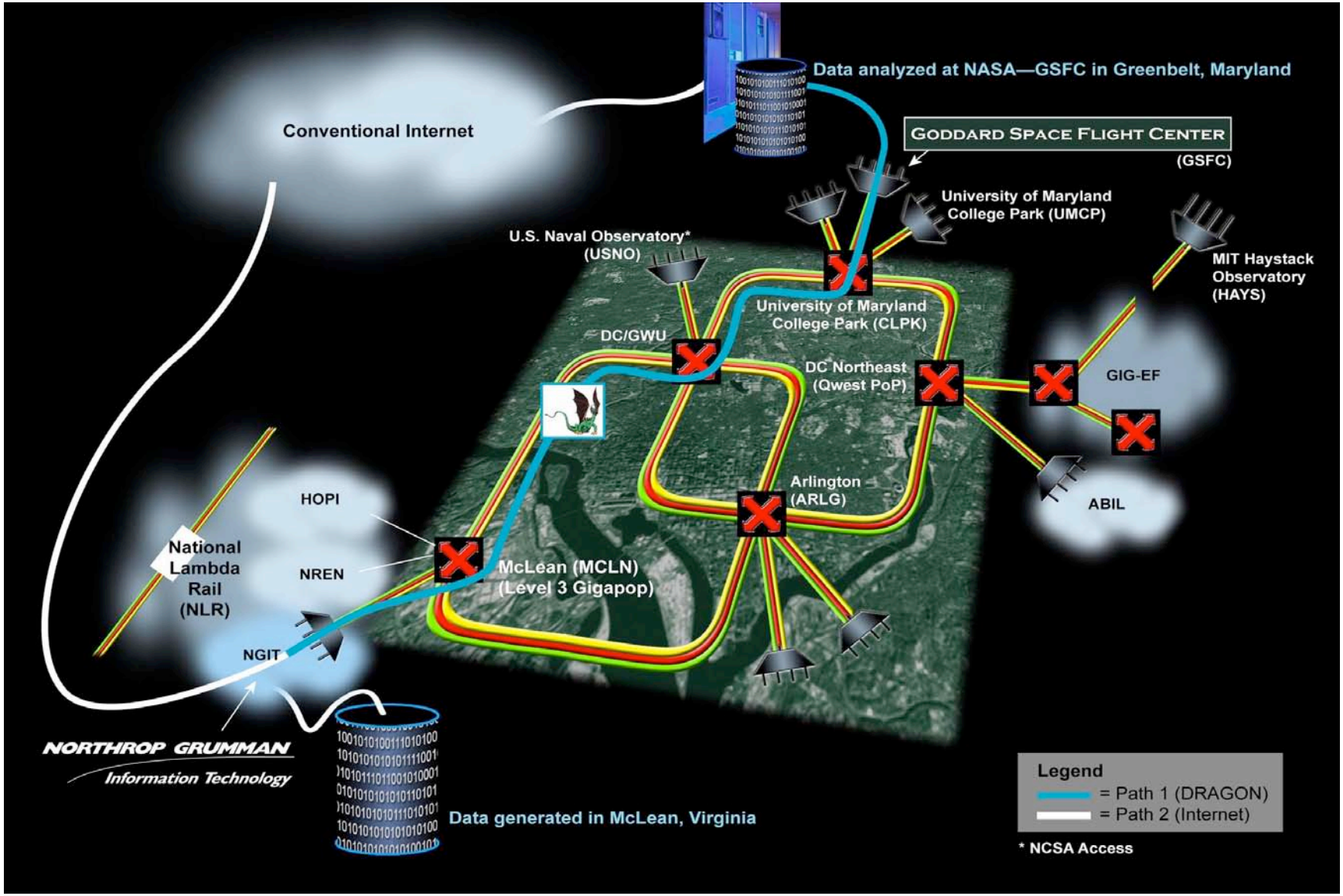
Courtesy Jerry Sobieski - jerrys@maxgigapop.net

Closing the loop: Hybrid Networks

- Hybrid Networks refer to emerging network technologies that allow the Internet to support both traditional *packet* based services as well as [new] *connection oriented* services.
- These services can coexist !
  - A user can [will] be able to access both from their workstation/cluster/lab/etc
  - IP services will likely run over and in conjunction with Light Path services, but other data formats are possible…
- These services will enable "affinity groups" to establish customized, dedicated, and highly dynamic network infrastructure that suits their needs
  - No longer will such specialized networks be expensive or complex
  - Such specialized networks will be able to evolve and morph to meet the changing needs of the collaborating organizations…

Courtesy Jerry Sobieski - jerrys@maxgigapop.net

The DRAGON's Lair - 3yr old baby picture



Courtesy Jerry Sobieski - jerrys@maxgigapop.net

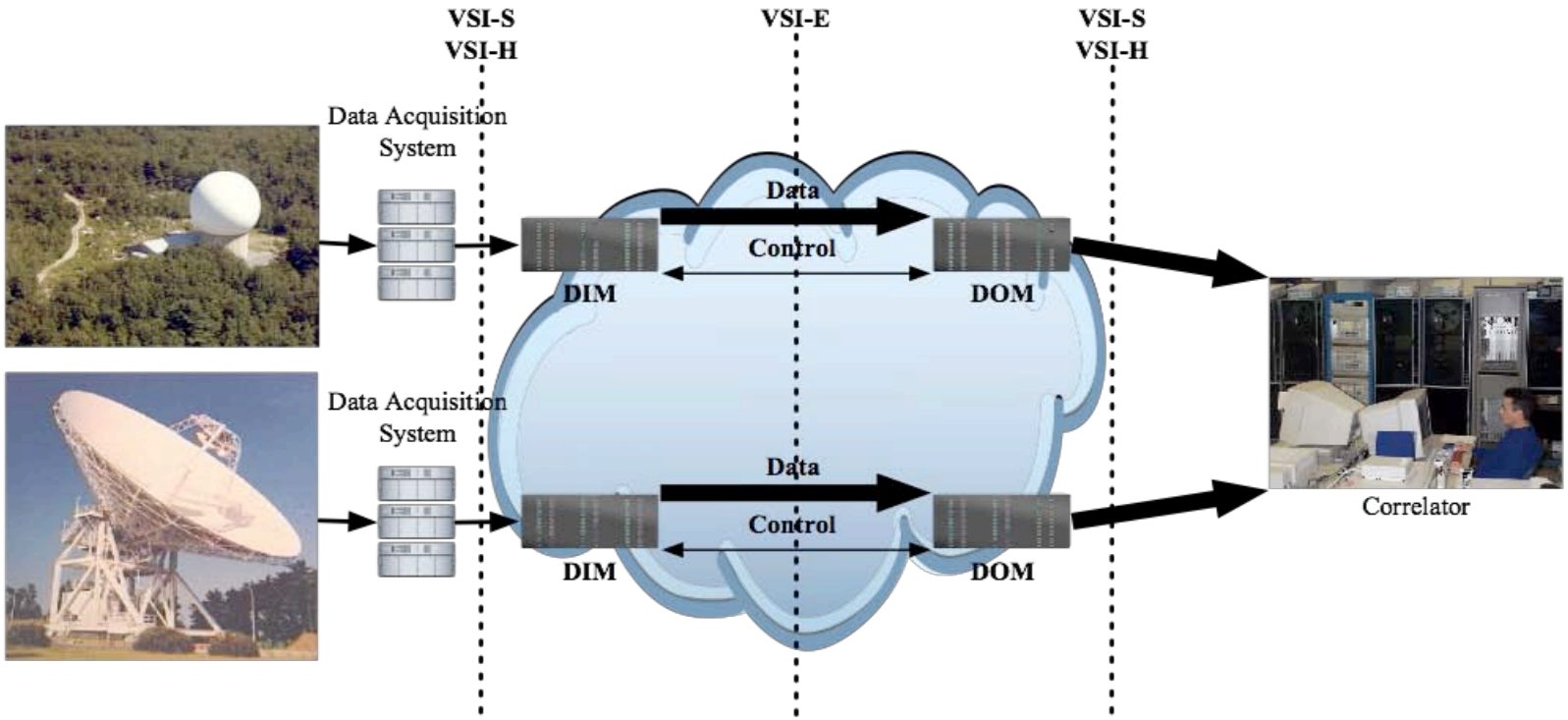
eVLBI will use RTP/RTCP

- eVLBI will use RTP to take advantage of the capabilities of that protocol.
- With the high data rate, the worry is that the time stamp may wrap too fast.
- So, the Haystack group proposes
  - A time stamp scaling bit and
  - A time stamp scaling factor
- If the scaling bit is set, the time stamps are assumed to be scaled by the scaling factor.

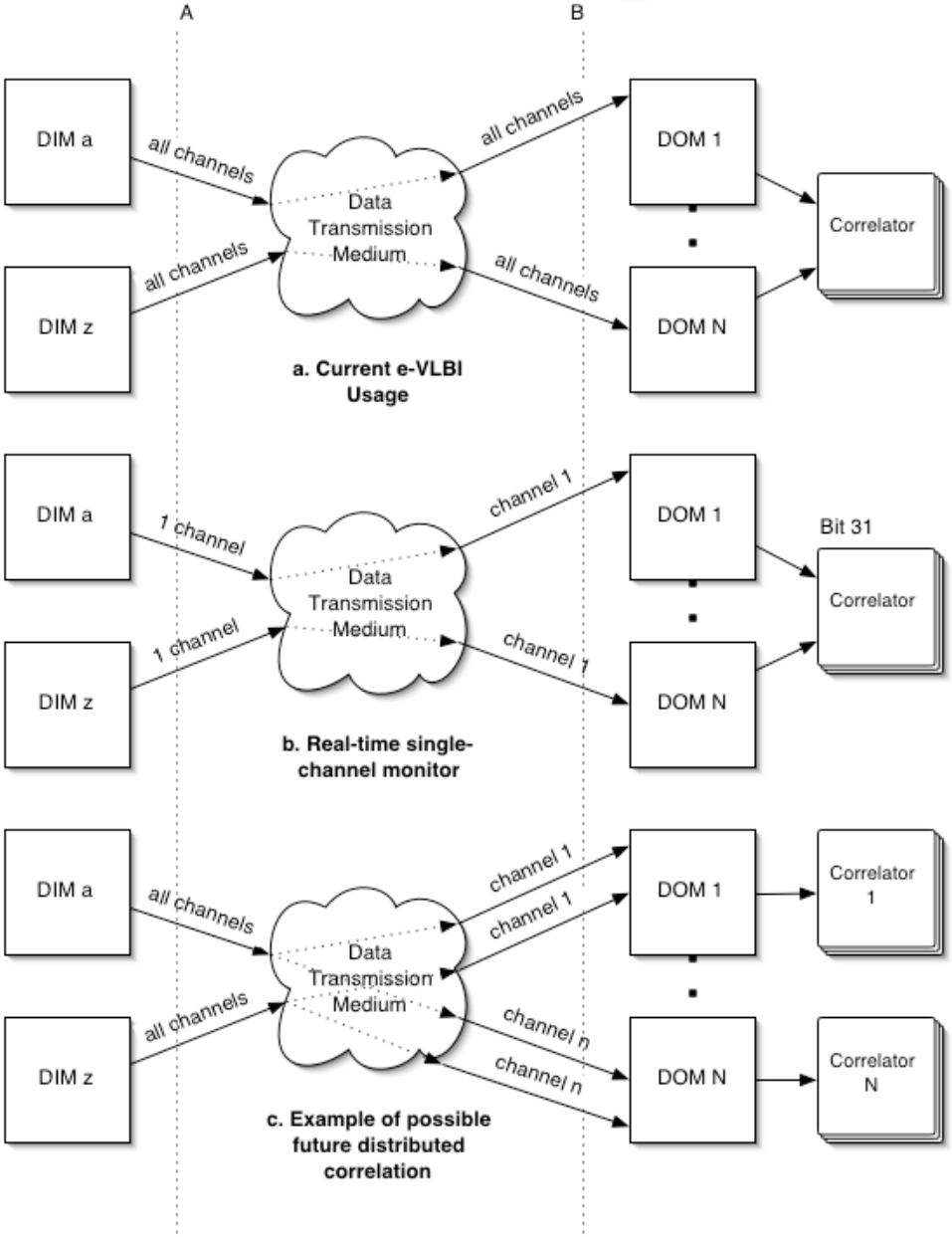
VLBI Standard Interface (VSI)

- VSI defines
  - A standard interface to and from a VLBI Data Transmission System (DTS)
  - Allows heterogeneous DTS's to be interfaced to both data-acquisition and correlator systems with a minimum of effort.
- VSI is defined to be compatible with:
  - tradition recording/playback systems,
  - network data transmission, and
  - direct-connect systems.

VSI's Model


VSI-S
VSI-H
VSI-E
VSI-S
VSI-H
Data Acquisition System
Data Acquisition System
Data
Control
DIM
DOM
Data
Control
DIM
DOM
Correlator

Network Topologies



VSI-E Proposal (cont)

- Why RTP/RTCP
  - RTP is the standard for real-time transport over IP
  - Transmission of sampled analog data
  - Dissemination of session information
  - Monitoring of network and end system performance (by participants and third parties)
  - Adaptation to varying network capability / performance
  - Appropriate reliability / repair model
  - Message Sequencing / un-reordering
  - Multi-cast distribution of statistics, control and data

RTP Extensions for e-VLBI

- RTP Profile for e-VLBI
  - defines the structure and semantics of the RTP packets used to transport VLBI data.
- Six packet types
  - RTP Data Packet
  - RTCP Sender Report Packet
  - RTCP Receiver Report Packet
  - RTCP Source DEScription Packet
  - RTCP BYE Packet
  - Application Defined RTCP Packet

Conclusion

- VSI-E
  - A media independent data format
    - Transmitted “on the wire”
  - Is compatible between **heterogeneous** DTSs
  - Efficient transport mechanism
  - Using Standard protocols
  - Internet-friendly transport
  - Scalable Implementation
  - Ability to transport individual data-channel streams as individual packet streams
  - Multicasting to transport data and/or control information in an efficient manner